

# ASYMPTOTIC THEORY OF INTEGRATED CONDITIONAL MOMENT TESTS<sup>1</sup>

By Herman J. Bierens<sup>2</sup> and Werner Ploberger<sup>3</sup>

In this paper we derive the asymptotic distribution of the test statistic of a generalized version of the integrated conditional moment (*ICM*) test of Bierens (1982, 1984), under a class of  $\sqrt{n}$ -local alternatives, where  $n$  is the sample size. The generalized version involved includes neural network tests as a special case, and allows for testing misspecification of dynamic models.

It appears that the *ICM* test has nontrivial local power. Moreover, we show that under the assumption of normal errors the *ICM* test is asymptotically admissible, in the sense that there does not exist a test that is uniformly more powerful.

The asymptotic size of the test is case-dependent: the critical values of the test depend on the data-generating process. In this paper we derive case-independent upperbounds of the critical values.

## 1. INTRODUCTION

Conditional moment (CM) tests have been proposed by Newey (1985) and Tauchen (1985) in the context of maximum likelihood models, but as these authors show, most misspecification tests of functional form are special forms of CM tests. A typical CM test takes the form of a quadratic form of finitely many weighted means of the residuals, where the weights are functions of the regressors. These CM tests are in general not consistent. In order to achieve consistency, Bierens' (1982, 1990) consistent conditional moment tests employ a class of weight functions indexed by a continuous nuisance parameter, so that actually uncountable many weight functions are employed. In order to obtain a single test statistic, Bierens (1982) proposes to integrate these nuisance parameters out. Therefore we shall call the test of Bierens (1982) the Integrated Conditional Moment (ICM) test. The test statistic of the CM test of Bierens (1990) is obtained by taking the supremum over the space of nuisance parameters.

In section 2 we review the ICM test and discuss the choice of the weight functions. In section 3 we derive the asymptotic distribution of the ICM test under a general class of  $\sqrt{n}$ -local alternatives, where we allow the data-generating process to be dependent. A  $\sqrt{n}$ -local alternative takes the form of an augmented regression model where the additional terms in the model vanish in probability at rate  $1/\sqrt{n}$ , where  $n$  is the sample size. It appears that the ICM test has nontrivial power against these  $\sqrt{n}$ -local alternatives. In section 4 we prove the admissibility of the ICM test, under the assumption of normal errors, i.e, we show that there does not exist a uniformly more powerful test.

Next to the conditional moment testing approaches of Bierens (1982, 1984, 1987, 1990), Bierens and Hartog (1988), De Jong (1995), De Jong and Bierens (1994), White (1989) and Stinchcombe and White (1991), there is also a competing line of recent literature on conditional moment tests based on comparison of parametric and (semi-)nonparametric models. See, e.g., Wooldridge (1992), Yatchew (1992), Gozalo (1993) and Hardle and Mammen (1993) for published papers in this area. Although not all of these authors derive local power results, the ones who do find local alternatives that shrink to the null at a slower rate than  $1/\sqrt{n}$ . Only Hardle and Mammen (1993) manage to achieve  $\sqrt{n}$ -local power, but only in one direction. In contrast, we will show in this paper that our ICM test has nontrivial  $\sqrt{n}$ -local power in all directions,

although the power is not the same in every direction.

Under the null hypothesis of model correctness, the test statistic of the ICM test is asymptotically distributed as an integral over a squared zero mean Gaussian process, where the covariance function of this Gaussian process depends on the distribution of the data and the functional form of the model. This makes it impossible to tabulate the exact asymptotic critical values of the ICM test. In section 5 we show how to derive upperbounds of the asymptotic critical values of the ICM test that are case-independent and can therefore be tabulated.

The proofs of theorems and lemmas are given in the appendix, except in cases where these proofs are also helpful in understanding the main argument. Also the assumptions (A and B) are stated in the appendix. Convergence results and conditions indicated by " $\rightarrow$ " that involve random variables refer to convergence in probability, unless otherwise stated. The indicator function is denoted by  $I(\cdot)$ , and indexed expectations signs, e.g.  $E_g$ , indicate that the expectation is taken under a certain hypothesis " $g$ ".

## 2. THE INTEGRATED CONDITIONAL MOMENT TEST

### 2.1. Introduction

Consider a random sample  $\{(y_t, x_t), t = 1, \dots, n\}$  from a  $k+1$ -variate distribution, or let  $(y_t, x_t)$  be a  $k+1$ -variate time series process, observable for  $t = 1, \dots, n$ , where  $y_t$  is the dependent variable and  $x_t$  is a  $k$ -vector of regressors (possibly containing lagged dependent variables). In parametric nonlinear regression analysis we usually specify the conditional expectation function of  $y_t$  relative to the vector  $x_t$  of regressors as a known function  $f(\cdot, \cdot)$  of  $x_t$  and a parameter vector  $\theta$ :

$$(1) \quad H_0: \exists \theta_0 \in \Theta \subset \mathbb{R}^m: P[E(y_t|x_t) = f(x_t, \theta_0)] = 1.$$

where  $\Theta$  is the parameter space. The consistent tests of Bierens (1982, 1990) test the null hypothesis (1) against the alternative:

$$(2) \quad H_1: \sup_{\theta \in \Theta} P[E(y_t|x_t) \neq f(x_t, \theta)] < 1.$$

Note that in the i.i.d. case the alternative (2) is just the complement of the null hypothesis (1), i.e., the alternative hypothesis involved is that the null hypothesis is false, but that in the time series case model correctness requires more conditions than only (1), namely the additional

condition that  $u_t = y_t - f(x_t, \theta_0)$  is a martingale difference sequence. The latter condition implies that  $E[u_t w_t] = 0$  for any function  $w_t$  of the past  $(y_{t-1}, x_{t-1}), (y_{t-2}, x_{t-2}), (y_{t-3}, x_{t-3}), \dots$  of the time series under review. The properties of the ICM test under data dependence is treated in different ways by Bierens (1984) and De Jong (1995), for the case of testing model correctness against all global alternatives. In this paper we also derive the asymptotic theory of the ICM under data-dependence, but now we test a parametric model against  $\sqrt{n}$ -local alternatives.

The idea behind the conditional moment test as introduced by Newey (1985) and Tauchen (1985) is to base a test statistic on a finite number of weighted mean of the estimated residuals, where the weights are functions of exogenous and lagged dependent variables (or instrumental variables). However, given a finite set of instruments, it is always possible to construct a data-generating process for which the null hypothesis is false but the power of the test is trivial. In order to have power against all deviations from the null hypothesis we need an infinite set of instruments, say  $w_t(\xi)$ , where  $\xi$  is contained in an index set  $\Xi$ . Now consider the random function  $\hat{z}(\xi) = (1/\sqrt{n}) \sum_{t=1}^n [y_t - f(x_t, \hat{\theta})] w_t(\xi)$ ,  $\xi \in \Xi$ . As is shown in Bierens (1990) for the i.i.d. case, under the null hypothesis this random function converges weakly to a continuous Gaussian random function  $z(\xi)$ , while under the alternative,  $\hat{z}(\xi)/\sqrt{n}$  converges to a nonstochastic nonzero limit function, for weight functions  $w_t(\xi) = \exp(\xi^T \Phi(x_t))$ , with  $\Phi$  a bounded one-to-one mapping. De Jong (1995) proves a similar result for time series models for the case where  $\Xi$  grows in dimension to infinity with the sample size. Again, in this paper we focus on the asymptotic theory of ICM tests under local alternatives, where the dimension of the compact set  $\Xi$  remains fixed.

The test statistic of the ICM test takes the form

$$(3) \quad \hat{T} = \int |\hat{z}(\xi)|^2 d\mu(\xi)$$

where  $\mu(\xi)$  is a probability measure on  $\Xi$ . This is (in essence) the form of the integrated consistent conditional moment test proposed by Bierens (1982).

The critical values of the ICM test are case-dependent. However, the asymptotic  $p$ -values can be consistently estimated, using the conditional Monte Carlo approach of Hansen (1990) and De Jong (1995). Denoting the estimated  $p$ -value involved by  $\tilde{p}$ , the ICM test is then applied in

the form of an asymptotic  $\alpha$ -level test

$$(4) \quad \tau_n = I(\tilde{p} < \alpha),$$

where  $\alpha$  is the significance level. Thus we reject the null hypothesis at the  $\alpha$  significance level if  $\tau_n = 1$ . Note that under the null hypothesis,  $E(\tau_n) \rightarrow \alpha$ . Only for the ICM test in this form we can show asymptotic admissibility, i.e., we shown that there does not exist an uniformly more powerful test.

## 2.2. The weight functions

The consistency of the ICM test (3) depends on the choice of the weight function  $w_i(\xi)$ . In Bierens (1990) it has been shown that the ICM test based on the weight function  $w_i(\xi) = \exp(\xi^T \Phi(x_i))$ , with  $\Phi$  a bounded one-to-one mapping, is consistent. Earlier, Bierens (1982) showed the consistency of the ICM test for the complex-valued weight function  $w_i(\xi) = \exp(i\xi^T \Phi(x_i))$ , and  $\mu$  the Lebesgue measure. Stinchcombe and White (1991) show that these consistency results carry over to a much wider class of weight functions than only  $\exp(\cdot)$ . For example, we may replace  $\exp(u)$  by the logistic function  $1/(1+\exp(-u))$ , which then gives rise to White's (1989) neural network<sup>4</sup> version of the randomized CM tests of Bierens (1987, 1988, 1994b, Ch.5). See also Lee, White and Granger (1993). For the purpose of the ICM test, however, the following straightforward extension of Theorem 1 of Bierens (1982) is sufficiently general:

**THEOREM 1:** *Let  $u$  be a random variable satisfying  $E|u| < \infty$ , and let  $x$  be a bounded  $k$ -variate random vector such that  $P(E(u|x) = 0) < 1$ . If  $w(u)$  is a complex or real valued function that is infinitely many times continuously differentiable in  $u = 0$  and satisfies the condition*

$$(5) \quad \{s \in \mathbb{N}: (d/du)^s w(u)|_{u=0} = 0\} \text{ is finite,}$$

*then  $\forall \varepsilon > 0 \exists \xi \in \mathbb{R}^k: E[u \cdot w(\xi^T x)] \neq 0$  and  $\|\xi\| < \varepsilon$ .*

The result in Theorem 1 implies that if we choose the measure  $\mu$  such that a small open neighborhood of the origin of  $\Xi$  is in its support, and  $x$  and  $w$  are as in Theorem 1, then  $P(E(u|x) = 0) < 1$  if and only if  $\int [E[u \cdot w(\xi^T x)]]^2 d\mu(\xi) > 0$ . Note that if the vector  $x$  is not bounded, we can without loss of generality replace  $x$  in Theorem 1 by  $\Phi(x)$ , with  $\Phi$  a bounded one-to-one mapping,

for conditioning on  $x$  is equivalent to conditioning on  $\Phi(x)$ . Moreover, note that the exponential and logistic functions, as well as (e.g.) the weight function  $w(u) = \cos(u) + \sin(u)$ , all satisfy condition (5). In the sequel of this paper, however, we shall leave the type of the weight function open, apart from being real valued, as consistency of the ICM test is not the main issue of the present research.

### 3. THE LIMITING DISTRIBUTION OF THE ICM TEST UNDER LOCAL ALTERNATIVES AND DATA-DEPENDENCE

#### 3.1. *The null model, the local alternative, and maintained hypotheses*

In the sequel we shall suppress the vector  $x_t$  of regressors in the regression function  $f(x, \theta)$  and the weight function  $w_t$ , in order to allow for models with infinitely many lagged dependent variables  $y_{t-j}$  and lagged *exogenous* explanatory variables  $x_{t-j}$  ( $j = 1, 2, 3, \dots$ ) but finitely many parameters such as ARMA and ARMAX models, and to allow for a possible distinction between regressors and instrumental variables. Thus under the null hypothesis we reformulate the model as

$$(6) \quad H_0: y_t = f_t(\theta_0) + u_t, \quad \theta_0 \in \Theta,$$

and under the local alternative as

$$(7) \quad H_1^L: y_{t,n} = f_t(\theta_0) + g_t/\sqrt{n} + u_t,$$

where the error  $u_t$  are martingale differences. The detailed maintained hypotheses regarding the  $f_t$ ,  $g_t$  and the weight functions  $w_t(\xi)$  are given in the appendix, as Assumption A. These assumptions allow the  $g_t$ 's to depend on lagged dependent variables as well. However, in the presence of lagged dependent variables in  $f_t(\theta)$  and/or  $g_t$  there are two, possibly different, interpretations of the local alternative (7). The first interpretation is that the lagged dependent variables in  $f_t(\theta)$  and  $g_t$  are generated by the null model. Thus, the local alternative (7) is then actually of the form  $y_{t,n} = y_t + g_t/\sqrt{n}$ , where the  $y_t$ 's are generated by the null model (6). The second interpretation is that the lagged dependent variables in  $f_t$  and  $g_t$  are now the lagged  $y_{t,n}$  generated by (7). The latter interpretation makes the random variables  $f_t(\theta_0)$  and  $g_t$  triangular arrays. Although all our assumptions and proofs are stated in terms of single arrays, our results straightforwardly carry over to triangular arrays. The same applies to the weight functions  $w_t(\xi)$ .

Under the local alternative (7) the process  $\hat{z}(\xi)$  now becomes

$$(8) \quad \hat{z}(\xi) = \frac{1}{\sqrt{n}} \sum_{t=1}^n [u_t + g_t/\sqrt{n} + f_t(\theta_0) - f_t(\hat{\theta})] w_t(\xi).$$

where  $\hat{\theta}$  is the nonlinear least squares estimator of  $\theta_0$ . Then it follows from (8) that under Assumption A, similarly to Bierens (1990),

$$(9) \quad \hat{z}(\xi) = \frac{1}{\sqrt{n}} \sum_{t=1}^n u_t \phi_t(\xi) + \frac{1}{n} \sum_{t=1}^n g_t \phi_t(\xi) + o_p(1),$$

uniformly over  $\xi$  in  $\Xi$ , where

$$(10) \quad \phi_t(\xi) = w_t(\xi) - b(\theta_0, \xi)^T A(\theta_0)^{-1} (\partial/\partial \theta^T) f_t(\theta_0),$$

with

$$A(\theta) = \text{plim}_{n \rightarrow \infty} (1/n) \sum_{t=1}^n \{(\partial/\partial \theta^T) f_t(\theta)\} \{(\partial/\partial \theta) f_t(\theta)\},$$

$$b(\theta, \xi) = \text{plim}_{n \rightarrow \infty} (1/n) \sum_{t=1}^n (\partial/\partial \theta^T) f_t(\theta) w_t(\xi)$$

Denoting

$$(11) \quad z_n(\xi) = (1/\sqrt{n}) \sum_{t=1}^n u_t \phi_t(\xi) + (1/n) \sum_{t=1}^n g_t \phi_t(\xi),$$

we thus have that under Assumption A,

$$(12) \quad \text{plim}_{n \rightarrow \infty} \sup_{\xi \in \Xi} |\hat{z}(\xi) - z_n(\xi)| = 0.$$

### 3.2. The limiting distribution of the ICM test under local alternatives

Assumption A guarantees the tightness of the process  $z_n(\cdot)$  defined by (11) and the asymptotic normality of the finite distributions of  $z_n(\cdot)$ . See the appendix. Consequently, we have:

**THEOREM 2:** *Let Assumption A hold. If  $H_1^L$  is true then  $\hat{z} \Rightarrow z$ , where  $z$  is a Gaussian*

process on  $\Xi$  with mean function  $\eta(\xi) = \text{plim}_{n \rightarrow \infty} (1/n) \sum_{t=1}^n g_t \phi_t(\xi)$  and covariance function

$\Gamma(\xi_1, \xi_2) = \text{plim}_{n \rightarrow \infty} (1/n) \sum_{t=1}^n u_t^2 \phi_t(\xi_1) \phi_t(\xi_2)$ . Then by the continuous mapping theorem,

$$(13) \quad \hat{T} \rightarrow T = \int z^2(\xi) d\mu(\xi) \text{ in distr.}$$

In order to analyze the nature of the limiting distribution  $T$  in (13), we need the following version of Mercer's theorem and its corollary:

LEMMA 1: (Mercer's Theorem) Let  $\Gamma(\xi_1, \xi_2)$  be a real valued positive semi-definite continuous function on  $\Xi \times \Xi$ , where  $\Xi$  is a compact space, and let  $\mu$  be a probability measure on  $\Xi$ . The solutions  $\lambda_i$  and  $\psi_i(\cdot)$ ,  $i = 1, 2, 3, \dots$  of the Eigenvalue problem  $\int \Gamma(\xi_1, \xi_2) \psi_i(\xi_2) d\mu(\xi_2) = \lambda_i \psi_i(\xi_1)$  are real valued and the function  $\Gamma$  has the series representation  $\Gamma(\xi_1, \xi_2) = \sum_{i=1}^{\infty} \lambda_i \psi_i(\xi_1) \psi_i(\xi_2)$ , where the series involved converges uniformly on  $\Xi \times \Xi$ .

LEMMA 2: Let the conditions of Lemma 1 be satisfied. The Eigenvalues  $\lambda_i$  are nonnegative and satisfy  $\sum_{i=1}^{\infty} \lambda_i < \infty$ . Moreover, the Eigenfunctions  $\psi_i(\cdot)$  are continuous and can be chosen orthonormal and complete in the space  $C(\Xi)$  of continuous real functions on  $\Xi$  as well as on the space  $L_2(\mu)$  of squared integrable functions w.r.t.  $\mu$ , i.e.:  $\int \psi_i(\xi) \psi_j(\xi) d\mu(\xi) = I(i = j)$ , and every function  $\phi$  in  $C(\Xi)$  or  $L_2(\mu)$  can be written as

$$(14) \quad \phi(\xi) = \sum_{i=1}^{\infty} g_i \psi_i(\xi) \text{ a.s. } L_2(\mu),$$

with Fourier coefficients

$$(15) \quad g_i = \int \phi(\xi) \psi_i(\xi) d\mu(\xi)$$

satisfying  $\sum_{i=1}^{\infty} g_i^2 < \infty$ .



Now let the function  $\Gamma$  in Lemma 1 be equal to the limit function in Theorem 2. Note that the continuity of  $z()$  and the compactness of  $\Xi$  imply that  $z()$  is square-integrable:  $z \in L_2(\mu)$  a.s. Since the set  $\{\psi_i(\xi), i = 1, 2, 3, \dots\}$  of Eigenfunctions is complete we can therefore apply Parseval's equality and conclude from (14) and (15), with  $\phi$  replace by  $z$ , that  $T = \sum_{i=1}^{\infty} [\int z(\xi) \psi_i(\xi) d\mu(\xi)]^2$ . Moreover, the Gaussianity of  $z()$  implies that the Fourier coefficients

$$(16) \quad \int z(\xi) \psi_i(\xi) d\mu(\xi), \quad i = 1, 2, 3, \dots$$

are Gaussian too. Therefore, for the characterization of their joint distribution we only need to compute covariances and means. The covariances are:

$$\begin{aligned} E \{ \int [z(\xi) - \eta(\xi)] \psi_i(\xi) d\mu(\xi) \int [z(\xi) - \eta(\xi)] \psi_j(\xi) d\mu(\xi) \} \\ = \int \int \Gamma(\xi_1, \xi_2) \psi_i(\xi_1) \psi_j(\xi_2) d\mu(\xi_1) d\mu(\xi_2) = \lambda_i I(i=j), \end{aligned}$$

so that the sequence (16) is independent. Moreover, it is easy to see that the mean of the  $i$ -th element of the sequence (16) is just the  $i$ -th Fourier coefficient of  $\eta()$ :

$$(17) \quad \eta_i = \int \eta(\xi) \psi_i(\xi) d\mu(\xi).$$

Therefore, the asymptotic distribution of the ICM test under the local alternative (7) can be described as follows:

**THEOREM 3:** *Under the local alternative (7) and Assumption A,  $T = \int z(\xi)^2 d\mu(\xi) \sim \sum_{i=1}^{\infty} (\eta_i + \varepsilon_i \sqrt{\lambda_i})^2$ , where the  $\varepsilon_i$  are i.i.d.  $N(0,1)$ , and the  $\eta_i$  are defined by (17).*

Note that the Eigenvalues  $\lambda_i$  depend on the covariance function  $\Gamma$ , which in its turn depends on the data-generating process under the null. Cf. Bierens (1990). Therefore, the asymptotic null distribution

$$(18) \quad T_0 = \sum_{i=1}^{\infty} \varepsilon_i^2 \lambda_i,$$

where  $\varepsilon_i$  is i.i.d.  $N(0,1)$ , is case-dependent. Moreover, note that the result of Theorem 3 implies

that in general the ICM test has nontrivial  $\sqrt{n}$ -local power:

COROLLARY 1: *If the mean function  $\eta(\xi)$  is such that*

$$(19) \quad \int \eta(\xi)^2 d\mu(\xi) > 0.$$

*then for every  $K > 0$ ,  $P(T > K) > P(T_0 > K)$ .*

Condition (19) can be achieved by a suitable choice of the weight functions  $w_i(\xi)$  and the measure  $\mu(\xi)$ . Cf. Section 2.

## 4. ADMISSIBILITY OF THE ICM TEST

### 4.1. Introduction

We show now, by adapting the approach of Andrews and Ploberger (1993, 1994), that the ICM test is asymptotically admissible, i.e., that there does not exist a test which uniformly dominates the asymptotic local power of the ICM test, provided the errors  $u_t$  are conditionally normally distributed and some regularity conditions hold. See Assumption B in the appendix.

Consider probability measures  $P_{0,n}$ , the probability measures which generate the data under the null hypothesis, and a family of probability measures  $P_{g,n}$ ,  $g \in G$ , representing alternatives. One may interpret the index  $g$  as the functional form of the random variables  $g_t$  in model (7), i.e.  $g_t = g(y_{t-1}, y_{t-2}, \dots, x_t, x_{t-1}, x_{t-2}, \dots)$ . In particular, we confine the index set  $G$  of alternatives to local alternatives (7) for which Assumption B holds. Note that for such an alternative  $g$  we can define  $P_{g,n}$  indirectly by the likelihood ratio  $dP_{g,n}/dP_{0,n}$ , which under Assumption B is well-defined, so that both  $P_{0,n}$  and  $P_{g,n}$  are defined on the same probability space.

Next, consider weighted alternatives  $P_{1,n} = \int P_{g,n} dQ_n(g)$ , where the  $Q_n$  are probability measures on  $G$ . The  $\alpha$ -level likelihood ratio test for testing  $P_{0,n}$  against  $P_{1,n}$  takes the form  $\rho_n = I(dP_{1,n}/dP_{0,n} > K_{\alpha,n})$ , where  $K_{\alpha,n}$  is the corresponding  $\alpha$ -fractile of the likelihood ratio involved. We shall show that under the null our ICM test  $\tau_n$  in the form (4) is asymptotically equivalent to the LR test for a particular measure  $Q_n$ , i.e.,  $P_{0,n}(\tau_n = \rho_n) \rightarrow 1$ . Now consider an arbitrary sequence  $\gamma_n$  of asymptotic  $\alpha$ -level tests competing with  $\tau_n$ . We distinguish three cases.

The first case is where  $\gamma_n$  and  $\tau_n$  are asymptotically equivalent under the null, i.e.,

$$(20) \quad P_{0,n}(\tau_n = \gamma_n) \rightarrow 1.$$

Then we can show that in the case of the ICM test these two tests are also equivalent under all alternatives  $P_{g,n}$ ,  $g \in G$ , i.e.,

$$(21) \quad P_{g,n}(\tau_n = \gamma_n) \rightarrow 1, \text{ for each } g \in G.$$

The second case is where  $\gamma_n$  and  $\tau_n$  are essentially different under the null, in the sense that

$$(22) \quad \liminf_{n \rightarrow \infty} P_{0,n}(\tau_n \neq \gamma_n) > 0.$$

Then we can show that

$$(23) \quad \liminf_{n \rightarrow \infty} \left( \int (E_g \tau_n) dQ_n(g) - \int (E_g \gamma_n) dQ_n(g) \right) > 0.$$

Thus, in this case the tests  $\tau_n$  have the highest "average" (w.r.t.  $Q_n$ ) asymptotic power. The third case is where neither (20) nor (22) are true. Then there exists a subsequence  $n_j$  along which the two tests are asymptotically equivalent under  $P_{0,n}$ , and thus also under all alternatives  $P_{g,n}$ .

In the first case the result (21) implies that the asymptotic power functions of the two tests are the same, hence  $\gamma_n$  cannot be asymptotically uniformly more powerful than  $\tau_n$ . The same applies in the third case, because we can approach the asymptotic power function along any subsequence. In the second case the result (23) implies that

$$\liminf_{n \rightarrow \infty} Q_n(\{g \in G: E_g \tau_n > E_g \gamma_n\}) > 0,$$

which excludes the possibility that asymptotically the test  $\gamma_n$  is uniformly more powerful than  $\tau_n$ .

For proving the result (21), we need:

**LEMMA 3:** *Let Assumption B hold. If the tests  $\gamma_n$  and  $\tau_n$  are asymptotically equivalent under  $P_{0,n}$  then so are they under  $P_{g,n}$ .*

Moreover, for proving (23) we need:

LEMMA 4: Let  $L_n = dP_{1,n}/dP_{0,n}$  be the likelihood ratio. Assume that under the null  $P_{0,n}$ ,  $L_n$  converges in distribution to a continuously distributed random variable  $L$  with  $E(L) = 1$ . If under  $P_{0,n}$  the asymptotic  $\alpha$ -level test  $\tau_n$  is asymptotically equivalent to the  $\alpha$ -level LR test  $\rho_n$ , and  $\gamma_n$  is a competing asymptotic  $\alpha$ -level test that is essentially different from  $\tau_n$ , i.e., (22) holds, then the asymptotic power of the test  $\tau_n$  is higher than the asymptotic power of the test  $\gamma_n$  (i.e., (23) holds).

The lemmas 3 and 4 are concerned with tests of simple hypotheses, whereas in the case of the ICM test we have composite hypotheses, because the null distribution as well as the alternative distribution depend on the parameter  $\theta_0$ . Thus, loosely speaking, the actual index set of alternatives is of the form  $G \times \Theta$ . However, this is no problem. If for all fixed  $\theta$  in the interior of  $\Theta$  there does not exist a test that is, uniformly on  $G$ , asymptotically more powerful than the ICM test, then there also cannot exist a test that is uniformly on  $G \times \Theta$  asymptotically more powerful than the ICM test, regardless of possible restrictions imposed on  $G$ . Therefore, we can now merge and extend the lemmas 3 and 4 to:

LEMMA 5: Let Assumption B hold, and let  $\tau_n$  be the ICM test in the form (4). Let  $L_{0,n}(\theta)$  be the likelihood of the data under the null hypothesis for a particular parameter vector  $\theta$  in  $\Theta$ . Similarly, let  $L_{1,n}(\theta, g)$  be the likelihood of the data under a particular alternative  $g \in G$  and a parameter vector  $\theta \in \Theta$ . Suppose that for any  $\theta$  in  $\Theta$  it is possible to construct probability measures  $Q_{\theta,n}$  which, with  $L_{1,n}(\theta) = \int L_{1,n}(\theta, g) dQ_{\theta,n}(g)$ , have the properties that under the null hypothesis,

$$(24) \quad \ln \left( \frac{L_{1,n}(\theta)}{L_{0,n}(\theta)} \right) - \frac{\hat{T}}{c} \rightarrow d(\theta)$$

and

$$(25) \quad \frac{L_{1,n}(\theta)}{L_{0,n}(\theta)} \rightarrow V_\theta \text{ in distr., where } E(V_\theta) = 1,$$

where  $c$  is a constant and  $d(\theta)$  a nonrandom function. Then the ICM test  $\tau_n$  is admissible.

Note that condition (24) ensures that the ICM test is asymptotically equivalent to a LR test, and that, since  $\hat{T}$  is asymptotically continuously distributed under the null, so is the likelihood ratio involved. Moreover, the conditions (24) and (25) ensure that the conclusion of Lemma 3 also holds for  $P_{1,n}$ ,

#### 4.2. Asymptotic admissibility

For proving the asymptotic admissibility of the ICM test it suffices now to construct probability measures  $P_g$  and  $Q_{\theta,n}$  that satisfy the conditions of Lemma 5, as follows. Denote

$$(26) \quad g_{t,i} = \int \phi_t(\xi) \psi_i(\xi) d\mu(\xi) \text{ if } t \geq 1, \quad g_{t,i,n} = 0 \text{ if } t < 1,$$

cf. (10) and (15). Then it follows from (12) that under the null hypothesis (6), with  $\theta (= \theta_0)$  any point in the parameter space  $\Theta$  satisfying Assumption A, that

$$\begin{aligned} \hat{T} &= \int z_n(\xi)^2 d\mu(\xi) + o_p(1) = \sum_{i=1}^{\infty} \left( \int z_n(\xi) \psi_i(\xi) d\mu(\xi) \right)^2 + o_p(1) \\ &= \sum_{i=1}^{\infty} \left( (1/\sqrt{n}) \sum_{t=1}^n (y_t - f_t(\theta)) \int \phi_t(\xi) \psi_i(\xi) d\mu(\xi) \right)^2 + o_p(1), \end{aligned}$$

hence under Assumption A and the null hypothesis,

$$(27) \quad \hat{T} = \sum_{i=1}^{\infty} \left( (1/\sqrt{n}) \sum_{t=1}^n (y_t - f_t(\theta)) g_{t,i} \right)^2 + o_p(1).$$

However, the random variables  $g_{t,i}$  in (26) also form the basis for the following class of alternative hypotheses:

$$(28) \quad H_1^L: y_{t,n} = f_t(\theta) + (\sigma/\sqrt{n}) \sum_{i=1}^{N_n} v_i g_{t,i} + u_t,$$

where the  $v_i$ 's are random coefficients and  $N_n$  converges to infinity with  $n$  at a sufficiently slow rate. We can associate these alternatives to a subset  $G_n$  of the set  $G$  of alternatives considered in Lemmas 3, 4 and 5. Thus, each alternative  $g$  in  $G_n$  corresponds to a sequence  $v_i$  of coefficients, with  $v_i = 0$  for  $i > N_n$ , and the null hypothesis corresponds to the case  $v_i = 0$  for  $i = 1, 2, \dots$

Given a  $g$  in  $G_n$ , we can now write the log likelihood ratio under the alternative  $g$  as:

$$\ln \left( \frac{L_{1,n}(\theta, g)}{L_{0,n}(\theta)} \right) = \frac{1}{\sigma \sqrt{n}} \sum_{i=1}^{N_n} v_i \sum_{t=1}^n g_{t,i}(y_t - f_t(\theta)) - \frac{1}{2} (1/n) \sum_{t=1}^n \left( \sum_{i=1}^{N_n} v_i g_{t,i} \right)^2.$$

Let  $g_N$  be an alternative for which  $v_i = 0$  for  $i > N$ , where  $N$  may possibly depend on  $n$ . Denoting

$$(29) \quad a_N = \frac{1}{\sigma} \left( (1/\sqrt{n}) \sum_{t=1}^n g_{t,1}(y_t - f_t(\theta)), \dots, (1/\sqrt{n}) \sum_{t=1}^n g_{t,N}(y_t - f_t(\theta)) \right)^T$$

$$B_N = \left( (1/n) \sum_{t=1}^n g_{t,i_1} g_{t,i_2} \right), \quad i_1, i_2 = 1, 2, \dots, N,$$

and  $V_N = (v_1, \dots, v_N)^T$ , we can now write the log likelihood ratio as

$$\ln \left( \frac{L_{1,n}(\theta, g_N)}{L_{0,n}(\theta)} \right) = a_N^T V_N - \frac{1}{2} V_N^T B_N V_N.$$

A suitable measure  $Q_n$  on  $G$  can now be constructed implicitly by letting  $V_N \sim N(0, (cI_N - \Lambda_N / \sigma^2)^{-1})$ , independently of the data, where  $\Lambda_N = \text{diag}(\lambda_1, \dots, \lambda_N)$  and  $c > \max\{\lambda_i / \sigma^2, i = 1, 2, 3, \dots\}$ , with  $N$  depending on  $n$ . Then

$$(30) \quad \frac{\int L_{1,n}(\theta, g) dQ_n(g)}{L_{0,n}(\theta)} = \frac{\sqrt{\det(cI_N - \Lambda_N / \sigma^2)} \exp\left(\frac{1}{2} a_N^T (B_N + (cI_N - \Lambda_N / \sigma^2))^{-1} a_N\right)}{\sqrt{\det(B_N + (cI_N - \Lambda_N / \sigma^2))}} \\ = L_{1,n}(\theta) / L_{0,n}(\theta), \text{ say.}$$

We show now that condition (24) of Lemma 5 holds. Observe from (10) and (15) that

$$\begin{aligned}
(1/n) \sum_{i=1}^n g_{t,i_1} g_{t,i_2} &= (1/n) \sum_{i=1}^n \iint (\phi_i(\xi_1) \phi_i(\xi_2) \psi_{i_1}(\xi_1) \psi_{i_2}(\xi_2) d\mu(\xi_1) d\mu(\xi_2)) \\
&= \frac{1}{\sigma^2} \iint \tilde{\Gamma}(\xi_1, \xi_2) \psi_{i_1}(\xi_1) \psi_{i_2}(\xi_2) d\mu(\xi_1) d\mu(\xi_2) \\
&\rightarrow \frac{1}{\sigma^2} \iint \Gamma(\xi_1, \xi_2) \psi_{i_1}(\xi_1) \psi_{i_2}(\xi_2) d\mu(\xi_1) d\mu(\xi_2) = \frac{\lambda_{i_1} I(i_1=i_2)}{\sigma^2}
\end{aligned}$$

where  $\tilde{\Gamma}(\xi_1, \xi_2) = (1/n) \sum_{i=1}^n E(u_i^2 | \mathcal{F}_{i-1}) \phi_i(\xi_1) \phi_i(\xi_2) = \sigma^2 (1/n) \sum_{i=1}^n \phi_i(\xi_1) \phi_i(\xi_2)$  and the convergence result involved follows from Assumptions A and B. Thus for fixed  $N$ ,  $B_N \rightarrow (1/\sigma^2) \Lambda_N$ . Consequently, it follows that under the null hypothesis and Assumptions A and B,  $a_N^T (B_N + cI_N - \Lambda_N/\sigma^2)^{-1} a_N - c^{-1} a_N^T a_N \rightarrow 0$  ( $n \rightarrow \infty, N$  fixed). Moreover, it follows from (27) and (29) that

$$\hat{T} - \sigma^2 a_N^T a_N = \sum_{i=N+1}^{\infty} \left( \int z_n(\xi) \psi_i(\xi) d\mu(\xi) \right)^2 + o_p(1).$$

Since for fixed  $N$ ,

$$\lim_{n \rightarrow \infty} E \left[ \sum_{i=N+1}^{\infty} \left( \int z_n(\xi) \psi_i(\xi) d\mu(\xi) \right)^2 \right] = \sum_{i=N+1}^{\infty} \lambda_i,$$

It follows now from Lemma 6 below that there exists a sequence  $N_n$  converging slowly to infinity with  $n$  such that:

$$(31) \quad a_{N_n}^T (B_{N_n} + cI_{N_n} - \Lambda_{N_n}/\sigma^2)^{-1} a_{N_n} - c^{-1} \hat{T}/\sigma^2 \rightarrow 0.$$

LEMMA 6: If  $A_{n,N}$  and  $B_N$  are random variables such that  $A_{n,N} \rightarrow B_N$  for fixed  $N$  and  $n \rightarrow \infty$ , and  $B_N \rightarrow 0$  for  $N \rightarrow \infty$ , then there exists a subsequence  $N_n^*$  converging to infinity with  $n$  such that  $A_{n,N_n} \rightarrow 0$  for all subsequences  $N_n$  satisfying  $N_n \leq N_n^*$  and  $N_n \rightarrow \infty$ .

Moreover, we have

$$(32) \quad \ln \left( \frac{\sqrt{\det(cI_N - \Lambda_N / \sigma^2)}}{\sqrt{\det(B_N + cI_N - \Lambda_N / \sigma^2)}} \right) \rightarrow \ln \left( \sqrt{\det(I_N - c^{-1} \Lambda_N / \sigma^2)} \right) \text{ a.s.} \\ (n \rightarrow \infty, N \text{ fixed})$$

and

$$(33) \quad \ln \left( \sqrt{\det(I_N - c^{-1} \Lambda_N / \sigma^2)} \right) = \frac{1}{2} \sum_{i=1}^N \ln \left( 1 - \frac{\lambda_i}{c \sigma^2} \right) \rightarrow \frac{1}{2} \sum_{i=1}^{\infty} \ln \left( 1 - \frac{\lambda_i}{c \sigma^2} \right)$$

as  $N \rightarrow \infty$ . Again applying Lemma 6, we can replace  $N$  in (32) and (33) by the same sequence  $N_n$  as before. Combining (31), (32) and (33) then yields:

$$(34) \quad \ln \left( \frac{L_{1,n}(\theta)}{L_{0,n}(\theta)} \right) - \left( \frac{\hat{T}}{2c\sigma^2} + \frac{1}{2} \sum_{i=1}^{\infty} \ln \left( 1 - \frac{\lambda_i}{c \sigma^2} \right) \right) \rightarrow 0$$

under the null hypothesis, where the likelihood ratio involved is defined by (30) with  $N$  replaced by  $N_n$ . This proves part (24) of Lemma 5, with  $d(\theta) = (1/2) \sum_{i=1}^{\infty} \ln[1 - (\lambda_i / c \sigma^2)]$ . Part (25) of Lemma 5 follows from Theorem 3 and (34), i.e., under the null hypothesis

$$\frac{\hat{T}}{2c\sigma^2} + \frac{1}{2} \sum_{i=1}^{\infty} \ln \left( 1 - \frac{\lambda_i}{c \sigma^2} \right) \rightarrow \frac{1}{2} \sum_{i=1}^{\infty} \varepsilon_i^2 \frac{\lambda_i}{c \sigma^2} + \frac{1}{2} \sum_{i=1}^{\infty} \ln \left( 1 - \frac{\lambda_i}{c \sigma^2} \right) = \ln(V_{\theta}),$$

say, in distr., where the  $\varepsilon_i$ 's are i.i.d.  $N(0,1)$ . Clearly,  $E(V_{\theta}) = 1$ . This completes the proof of the following theorem:

**THEOREM 4:** *Under Assumptions A and B, the ICM test in the form (4) is admissible*

## 5. THE SIZE OF THE ICM TEST

As mentioned before, the practical applicability of the ICM test is hampered by the fact that the limiting distribution of the test statistic under the null hypothesis is case-dependent and can therefore not be tabulated. A possible way to get around this problem is the conditional Monte Carlo approach of Hansen (1990) and De Jong (1995). However, this approach is computational intensive and therefore does not give quick answers. Therefore, we shall derive



case-independent upperbounds of the asymptotic critical values of the ICM test, on the basis of the following lemma

LEMMA 7: Let  $c_1, \dots, c_n$  be positive constants such that the equality  $(1/k) \sum_{i=1}^k c_i = (1/m) \sum_{j=1}^m c_j$  implies  $k = m$ , and let  $x_1, \dots, x_n$  be variables. Then the solution of the LP problem  $\max \sum_{j=1}^n c_j x_j$  s.t.  $x_1 \geq x_2 \geq \dots \geq x_n$ ,  $\sum_{j=1}^n x_j = 1$ , is of the form  $x_j = 1/m$  for  $j = 1, \dots, m$ ;  $x_j = 0$  for  $j = m+1, \dots, n$ .

It follows now from Theorems 2, 3 and Lemma 7:

THEOREM 5: Let  $\varepsilon_j$  be  $NID(0,1)$  and let

$$(35) \quad \bar{W} = \sup_{n \geq 1} (1/n) \sum_{j=1}^n \varepsilon_j^2.$$

For  $\eta > 0$ ,  $P[T_0 > \eta E(T_0)] \leq P[\bar{W} > \eta]$ , where  $T_0$  is the random variable defined in (18).

Consequently, under Assumption A and the null hypothesis (6),

$$\lim_{n \rightarrow \infty} P[\hat{T} > \eta \int \hat{\Gamma}(\xi, \xi) d\mu(\xi)] \leq P[\bar{W} > \eta].$$

Using 10,000 replications, we have derived the 10%, 5% and 1% quantiles of the random variable (35) by Monte Carlo simulation, i.e.,

$$(36) \quad P(\bar{W} > 3.23) = 0.10; P(\bar{W} > 4.26) = 0.05; P(\bar{W} > 6.81) = 0.01.$$

Thus, conducting the ICM test at say the 5% significance level, we reject the null hypothesis of model correctness if

$$(37) \quad \hat{T}_n > 4.26 \int \hat{\Gamma}(\xi, \xi) d\mu(\xi),$$

Note that Bierens (1982) proposed to derive critical values of the ICM test on the basis of Chebishev's inequality for first moments; e.g, under  $H_0$ ,

$$(38) \quad \lim_{n \rightarrow \infty} P[\hat{T}_n > 20 \int \hat{\Gamma}(\xi, \xi) d\mu(\xi)] \leq 0.05$$

Comparing (37) and (38) we see that the new upperbounds of the critical values in (36) are much sharper than the ones based on Chebishev's inequality.

*Southern Methodist University, Texas, USA & Tilburg University, the Netherlands,  
and  
Technische Universität Wien, Austria*

*First draft: March 1994.*

*Present revision: June 1995*

## APPENDIX

### ASSUMPTIONS

ASSUMPTION A.1: *The parameter space  $\Theta$  is a compact subset of  $\mathbb{R}^n$ . The true parameter vector  $\theta_0$  is contained in the interior of  $\Theta$ . The response function  $f_t(\theta)$  is twice continuously differentiable on  $\Theta$ , and  $u_t$  and  $f_{t+1}(\theta)$  are measurable w.r.t.  $\mathcal{F}_t$ , where  $\mathcal{F}_t$  is the sequence of  $\sigma$ -algebras generated by  $(y_{t-j}, x_{t-j})$ ,  $j = 0, 1, 2, \dots$ . Moreover,  $E(u_t | \mathcal{F}_{t-1}) = 0$  a.s. Furthermore,  $g_t$  is measurable w.r.t.  $\mathcal{F}_{t-1}$ ;  $g_t = 0$  for  $t < 1$ .*

ASSUMPTION A.2:  *$w_t(\xi)$  is a sequence of real valued random functions on  $\Xi$ , where  $\Xi$  is a compact subset of a Euclidean space, such that  $w_t(\xi)$  is measurable w.r.t.  $\mathcal{F}_{t-1}$ .*

ASSUMPTION A.3: *Let  $A_n(\theta) = (1/n) \sum_{t=1}^n \{(\partial/\partial\theta^T)f_t(\theta)\} \{(\partial/\partial\theta)f_t(\theta)\}$ . Then  $A_n(\theta) \rightarrow A(\theta)$  uniformly on  $\Theta$ , where  $A(\theta)$  is a nonstochastic matrix function such that  $A(\theta_0)$  is positive definite. Moreover, the least squares estimator  $\hat{\theta}$  satisfies*

$$\sqrt{n}(\hat{\theta} - \theta_0) = A(\theta_0)^{-1} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n u_t \frac{\partial}{\partial\theta^T} f_t(\theta_0) + \frac{1}{n} \sum_{t=1}^n g_t \frac{\partial}{\partial\theta^T} f_t(\theta_0) \right) + o_p(1),$$

ASSUMPTION A.4: *Let  $\hat{b}(\theta, \xi) = (1/n) \sum_{t=1}^n (\partial/\partial\theta^T)f_t(\theta)w_t(\xi)$ . Then  $\hat{b}(\theta, \xi) \rightarrow b(\theta, \xi)$  uniformly on  $\Theta \times \Xi$ , where  $b(\theta, \xi)$  is a nonstochastic function satisfying  $\sup_{\theta \in \Theta, \xi \in \Xi} \|b(\theta, \xi)\| < \infty$ .*

ASSUMPTION A.5: *The weight functions  $w_t(\xi)$  are differentiable on  $\Xi$ , and  $\limsup_{n \rightarrow \infty} (1/n) \sum_{t=1}^n E[u_t^2 \sup_{\xi \in \Xi} \|(\partial/\partial\xi^T)w_t(\xi)\|^2] < \infty$ ;  $(\partial/\partial\xi^T)\hat{b}(\theta, \xi) \rightarrow (\partial/\partial\xi^T)b(\theta, \xi)$  uniformly on  $\Theta \times \Xi$ ,  $\sup_{\theta \in \Theta, \xi \in \Xi} \|(\partial/\partial\xi^T)b(\theta, \xi)\| < \infty$ ;  $(1/n) \sum_{t=1}^n E\{u_t^2 [(\partial/\partial\theta^T)f_t(\theta_0)] \times [(\partial/\partial\theta)f_t(\theta_0)]\} \rightarrow A_2$  where  $A_2$  is finite. There exists a continuous function  $\Gamma(\xi_1, \xi_2)$  on  $\Xi \times \Xi$  such that  $(1/n) \sum_{t=1}^n E(u_t^2 | \mathcal{F}_{t-1}) \phi_t(\xi_1) \phi_t(\xi_2) \rightarrow \Gamma(\xi_1, \xi_2)$  uniformly on  $\Xi \times \Xi$ , while pointwise on  $\Xi \times \Xi$ ,*

$(1/n)\sum_{t=1}^n u_t^2 \phi_t(\xi_1)\phi_t(\xi_2) \rightarrow \Gamma(\xi_1, \xi_2)$ ,  $(1/n)\sum_{t=1}^n E[u_t^2 \phi_t(\xi_1)\phi_t(\xi_2)] \rightarrow \Gamma(\xi_1, \xi_2)$ . Moreover, for some  $\delta > 0$ ,  $\limsup_{n \rightarrow \infty} \sup_{\xi \in \Xi} (1/n)\sum_{t=1}^n E|u_t \phi_t(\xi)|^{2+\delta} < \infty$ . There exists a continuous function  $\eta(\xi)$  on  $\Xi$  such that  $(1/n)\sum_{t=1}^n g_t \phi_t(\xi) \rightarrow \eta(\xi)$  uniformly on  $\Xi$ .

**ASSUMPTION B:** The errors  $u_t$ 's in the models (6) and (7) are normally distributed:  $u_t | \mathcal{F}_{t-1} \sim N(0, \sigma^2)$ . Moreover, the exogenous variables  $x_t$ 's (c.f. Assumption A.1) are weakly exogenous in the sense of Engle, Hendry and Richard (1983). Furthermore, under the null hypothesis,  $\text{plim}_{n \rightarrow \infty} (1/n)\sum_{t=1}^n g_t^2$  exists, is constant and finite.

## PROOFS

**PROOF OF THEOREM 1:** The proof is similar to the related results in Bierens (1982, 1990).

**PROOF OF THEOREM 2:** We need to show that the finite distributions of the process  $z_n$  converge to normal distributions, and that  $z_n$  is tight. Cf. Billingsley (1968). In order to prove that the finite distributions of  $z_n$  are asymptotically normal, we apply the Liapounov-type version in Bierens (1994b, Th.6.1.7) of McLeish's (1974) martingale difference central limit theorem. Since pointwise in  $\xi$ ,  $u_t \phi_t(\xi)$  is a martingale difference sequence, Assumption A.5 implies that for arbitrary points  $\xi_1, \xi_2, \dots, \xi_M$  in  $\Xi$ ,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n u_t \begin{pmatrix} \phi_t(\xi_1) \\ \vdots \\ \phi_t(\xi_M) \end{pmatrix} \rightarrow N_M \left[ \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \Gamma(\xi_1, \xi_1) & \dots & \Gamma(\xi_1, \xi_M) \\ \vdots & \ddots & \vdots \\ \Gamma(\xi_M, \xi_1) & \dots & \Gamma(\xi_M, \xi_M) \end{pmatrix} \right]$$

in distr. Together with Assumption A.5, this result implies that the finite distributions of  $z_n$  converge to normal distributions.

Next, consider the following general tightness result:

**LEMMA A.1:** Let  $u_t$  be a martingale difference sequence, i.e.,  $u_t$  is measurable w.r.t.

$\mathcal{F}_t$ , where  $\mathcal{F}_t$  is an increasing sequence of  $\sigma$ -algebras and  $E(u_t | \mathcal{F}_{t-1}) = 0$  a.s. Moreover, let  $\phi_t(\xi)$  be a sequence of random function on a compact subset  $\Xi$  of a Euclidean space such that  $|\phi_t(\xi_1) - \phi_t(\xi_2)| \leq K_t \|\xi_1 - \xi_2\|$  for each  $\xi_1, \xi_2$  in  $\Xi$ , where  $\phi_t(\xi)$  and  $K_t$  are measurable w.r.t.  $\mathcal{F}_{t-1}$  and  $\limsup_{n \rightarrow \infty} (1/n) \sum_{t=1}^n E[u_t^2 K_t^2] < \infty$ . Finally, let for one arbitrary  $\xi_0$  in  $\Xi$ ,  $\limsup_{n \rightarrow \infty} (1/n) \sum_{t=1}^n E[u_t^2 \phi_t(\xi_0)^2] < \infty$ . Then the sequence of random functions  $z_n(\xi) = (1/\sqrt{n}) \sum_{t=1}^n u_t \phi_t(\xi)$  is tight on  $\Xi$ .

PROOF: Choose an arbitrary  $\varepsilon > 0$ . We prove the lemma by showing the existence of a sequence of tight random functions  $v_n(\xi)$  on  $\Xi$  such that  $P[z_n = v_n] \geq 1 - \varepsilon$ . Denote  $A_t(\xi) = \sum_{j=1}^t u_j^2 \phi_j(\xi)^2$ ,  $B_t = \sum_{j=1}^t u_j^2 K_j^2$ . Now choose an  $M > 0$  and define the stopping time  $\tau(M)$  by  $\tau(M) = \sup\{t \leq n \mid A_t(\xi_0) \leq nM, B_t \leq nM\}$  for an arbitrary  $\xi_0$  in  $\Xi$ . Since  $A_t(\xi_0)$  and  $B_t$  are monotonic non-decreasing, and  $\limsup_{n \rightarrow \infty} (1/n) E[A_n(\xi_0)] < \infty$ ,  $\limsup_{n \rightarrow \infty} (1/n) E[B_n] < \infty$  by the conditions of the lemma under review, it follows from Chebishev's inequality applied to  $A_n(\xi_0)$  and  $B_n$  that there exists an  $M_\varepsilon$  such that  $P[\tau(M_\varepsilon) = n] \geq 1 - \varepsilon$ . Next, define  $v_n(\xi) = z_{\tau(M_\varepsilon)}(\xi)$ . Then  $P[z_n = v_n] \geq P[\tau(M_\varepsilon) = n] \geq 1 - \varepsilon$ . We show that  $v_n$  is tight by applying the Kolmogorov-Cencov criterion (c.f. Kunita 1990, Theorem 1.4.7, p.38), i.e., if for some  $\gamma, \delta > 0$  there exists a constant  $C$  such that for every  $\xi_0, \xi_1, \xi_2$  in  $\Xi$ ,  $E|v_n(\xi_0)|^\gamma \leq C$ , and  $E|v_n(\xi_1) - v_n(\xi_2)|^\gamma \leq C \|\xi_1 - \xi_2\|^{k+\delta}$ , where  $k$  is the dimension of  $\Xi$ , then  $v_n$  is tight. Now utilize Burkholder's inequality (c.f. Chow and Teicher 1988, p. 396), i.e., if  $f_n$  is a martingale and  $S_n = \sum_{i=1}^n (f_i - f_{i-1})^2$ , then for  $m, n \geq 1$ ,  $|E[f_n^m]| \leq C_m E[S_n^{m/2}]$ , where  $C_m < \infty$  is a constant which does not depend on  $n$ . Moreover,  $n$  can be an arbitrary adapted and bounded stopping time. Applying this inequality to  $v_n$  yields

$$E |v_n(\xi_0)|^{2k+2} \leq C_{2k+2} (1/n^{k+1}) E \left( \sum_{t=1}^{\tau(M)} u_t^2 \phi_t(\xi_0)^2 \right)^{k+1} \leq C_{2k+2} M^{k+1},$$

where the second inequality follows from the definition of the stopping time  $\tau(M)$ . This proves the first part of the Kolmogorov-Cencov criterion, for  $\gamma = 2k + 2$ .

Finally, again using Burkholder's inequality, the Lipschitz condition on  $\phi_t$  and the definition of  $\tau(M)$ , it follows that

$$\begin{aligned} E |v_n(\xi_1) - v_n(\xi_2)|^{2k+2} &= (1/n^{k+1}) E \left( \sum_{i=1}^{\tau(M)} u_i (\phi_i(\xi_1) - \phi_i(\xi_2)) \right)^{2k+2} \\ &\leq C_k (1/n^{k+1}) E \left( \sum_{i=1}^{\tau(M)} u_i^2 (\phi_i(\xi_1) - \phi_i(\xi_2))^2 \right)^{k+1} \\ &\leq C_k (1/n^{k+1}) E \left( \sum_{i=1}^{\tau(M)} u_i^2 K_i^2 \right)^{k+1} \cdot \|\xi_1 - \xi_2\|^{2k+2} \leq C_k \|\xi_1 - \xi_2\|^{2k+2} M^{k+1}. \end{aligned}$$

This result proves the second part of the Kolmogorov-Cencov criterion and hence the tightness of  $v_n$ . Q.E.D.

In our case it follows from Assumption A.5 that we can choose  $K_i = \sup_{\xi \in \Xi} \|(\partial/\partial \xi^T) \phi_i(\xi)\|$ . Then all the conditions of Lemma A.1 follow easily from Assumption A. This completes the proof of Theorem 3.

PROOF OF COROLLARY 1: Denoting  $T_i = \sum_{j \neq i} (\eta_j + \varepsilon_j \sqrt{\lambda_j})^2$ , the corollary follows from repeated application of the easy inequality

$$\begin{aligned}
P(T \leq K) &= E \left( \int_{-(K-T_i)^{1/2}}^{(K-T_i)^{1/2}} \frac{\exp[-\frac{1}{2}(u-\eta_i)^2/\lambda_i]}{\sqrt{2\pi\lambda_i}} du I(T_i < K) \right) \\
&< E \left( \int_{-(K-T_i)^{1/2}}^{(K-T_i)^{1/2}} \frac{\exp[-\frac{1}{2}u^2/\lambda_i]}{\sqrt{2\pi\lambda_i}} du I(T_i < K) \right) \\
&= P(\lambda_i \varepsilon_i^2 + T_i \leq K) \text{ if } \eta_i \neq 0.
\end{aligned}$$

Thus if at least one  $\eta_i \neq 0$ , then the conclusion of the corollary holds. It is easy to verify that condition (19) guarantees this.

PROOF OF LEMMA 1: The series representation for  $\Gamma$  is the actual contents of Mercer's theorem. Cf. Dunford and Schwartz (1963, p.1088). The claim that the Eigenvalues and Eigenfunctions are real valued follows easily from the condition that the function  $\Gamma$  is real valued and positive semi-definite, similarly to the proof that the Eigenvalues and Eigenvectors of a positive semi-definite matrix are real valued.

PROOF OF LEMMA 2: The proof that the Eigenfunctions can be chosen orthonormal is analogous to the matrix case. The nonnegativity and summability of the Eigenvalues and the continuity of the Eigenfunctions follow directly from Mercer's theorem and the continuity of  $\Gamma$ . The completeness of the Eigenfunctions (part (14) of Lemma 2) follows from the fact that we can always make the orthonormal basis  $\{\psi_i\}$  complete by adding additional orthonormal functions with corresponding zero Eigenvalues.

PROOF OF LEMMA 3: Observe that

$$\begin{aligned}
P_{g,n}(\tau_n \neq \gamma_n) &\leq P_{g,n}(\{L_n > m\} \cup \{L_n \leq m \wedge \tau_n \neq \gamma_n\}) \\
&\leq P_{g,n}(\{L_n > m\}) + P_{g,n}(\{L_n \leq m \wedge \tau_n \neq \gamma_n\}) \\
&\leq \int_{\{L_n > m\}} L_n dP_{0,n} + \int_{\{L_n \leq m \wedge \tau_n \neq \gamma_n\}} L_n dP_{g,n} \leq \int_{\{L_n > m\}} L_n dP_n + m P_{0,n}(\tau_n \neq \gamma_n).
\end{aligned}$$

Thus for arbitrary  $m$  we have

$$(A1) \quad \limsup_{n \rightarrow \infty} P_{g,n}(\tau_n \neq \gamma_n) \leq \limsup_{n \rightarrow \infty} \int_{\{L_n > m\}} L_n dP_{g,n}.$$

Now if under  $P_{0,n}$ ,  $L_n \rightarrow L$  in distr., where  $L$  is a continuously distributed random variable satisfying  $E(L) = 1$ , then it follows from Lemma 6.12 in Strasser (1985, p.36) that  $L_n$  is uniformly  $(P_{g,n})$ -integrable and that therefore, by increasing  $m$ , we can make the right-hand side of (A1) arbitrarily small. But due to Assumption B we have under  $P_{0,n}$ ,

$$\ln(L_n) = \frac{-1}{\sigma^2} \left( (1/\sqrt{n}) \sum_{t=1}^n u_t g_t + (1/n) \sum_{t=1}^n g_t^2 \right) \rightarrow N(-\omega^2/2, \omega^2)$$

in distr., where  $\omega^2 = \text{plim}_{n \rightarrow \infty} (1/n) \sum_{t=1}^n g_t^2 / \sigma^2$ , hence  $L = \exp(-1/2\omega^2) \exp(\omega\varepsilon)$  with  $\varepsilon \sim N(0,1)$ , and obviously  $E(L) = 1$ .

PROOF OF LEMMA 4: Suppose first that the competing test  $\gamma_n$  is an exact  $\alpha$ -level test, and that  $\tau_n$  is an exact  $\alpha$ -level LR test:  $\tau_n = \rho_n$ . Let  $K_{\alpha,n}$  be the corresponding  $\alpha$ -fractile of the likelihood ratio  $L_n$ . Then we can write

$$\begin{aligned} \{\gamma_n \neq \rho_n\} &= \{\{\gamma_n \neq \rho_n\} \cap \{L_n < K_{\alpha,n}\}\} \cup \{\{\gamma_n \neq \rho_n\} \cap \{L_n \geq K_{\alpha,n}\}\} = \\ &\{\{\gamma_n = 1 \wedge \rho_n = 0\} \cap \{L_n < K_{\alpha,n}\}\} \cup \{\{\gamma_n = 0 \wedge \rho_n = 1\} \cap \{L_n \geq K_{\alpha,n}\}\}, \end{aligned}$$

hence

$$\gamma_n - \rho_n = I(\{\gamma_n \neq \rho_n\} \cap \{L_n < K_{\alpha,n}\}) - I(\{\gamma_n \neq \rho_n\} \cap \{L_n \geq K_{\alpha,n}\}).$$

Since under the null,  $E(\gamma_n) = E(\rho_n) = \alpha$ , these two equalities imply that

$$(A2) \quad P_{0,n}(\{\gamma_n \neq \rho_n\} \cap \{L_n < K_{\alpha,n}\}) = \frac{1}{2} P_{0,n}(\gamma_n \neq \rho_n).$$

Since we have assumed that the tests  $\gamma_n$  and  $\rho_n$  are essentially different, the "liminf" of the right-hand side probability is bounded away from zero, hence

$$(A3) \quad \liminf_{n \rightarrow \infty} P_{0,n}(\{\gamma_n \neq \rho_n\} \cap \{L_n < K_{\alpha,n}\}) > 0.$$

This result implies that there exists a  $\delta_0 > 0$  such that



$$(A4) \quad \liminf_{n \rightarrow \infty} P_{0,n}(\{\gamma_n \neq \rho_n\} \cap \{L_n < K_{\alpha,n} - \delta\}) > 0 \text{ if } 0 \leq \delta < \delta_0,$$

because if not then

$$(A5) \quad \begin{aligned} & \liminf_{n \rightarrow \infty} P_{0,n}(\{\gamma_n \neq \rho_n\} \cap \{L_n < K_{\alpha,n}\}) \\ & \leq \liminf_{n \rightarrow \infty} P_{0,n}(\{\gamma_n \neq \rho_n\} \cap \{L_n < K_{\alpha,n} - \delta\}) \\ & \quad + \liminf_{n \rightarrow \infty} P_{0,n}(K_{\alpha,n} - \delta \leq L_n < K_{\alpha,n}) \\ & = \inf_{\delta > 0} \liminf_{n \rightarrow \infty} P_{0,n}(K_{\alpha,n} - \delta \leq L_n < K_{\alpha,n}) = 0, \end{aligned}$$

which contradicts (A3).

Next, observe that

$$(A6) \quad \begin{aligned} E_1(\gamma_n) &= \int \gamma_n dP_{1,n} = \int \gamma_n L_n dP_{0,n} = E_0(\gamma_n L_n) \\ &= E_0[(L_n - K_{\alpha,n}) \gamma_n] + \alpha K_{\alpha,n}, \end{aligned}$$

and similarly  $E_1(\rho_n) = E_0[(L_n - K_{\alpha,n}) \rho_n] + \alpha K_{\alpha,n}$ . Thus

$$(A7) \quad \begin{aligned} E_1(\rho_n) - E_1(\gamma_n) &= E_0[(L_n - K_{\alpha,n})(\rho_n - \gamma_n)] \\ &= E_0[(L_n - K_{\alpha,n})(\rho_n - \gamma_n)I(L_n < K_{\alpha,n} - \delta)] \\ &\quad + E_0[(L_n - K_{\alpha,n})(\rho_n - \gamma_n)I(K_{\alpha,n} - \delta \leq L_n < K_{\alpha,n})] \\ &\quad + E_0[(L_n - K_{\alpha,n})(\rho_n - \gamma_n)I(L_n \geq K_{\alpha,n})] \\ &\geq E_0[(L_n - K_{\alpha,n})(\rho_n - \gamma_n)I(L_n < K_{\alpha,n} - \delta)] \\ &\quad + E_0[(L_n - K_{\alpha,n})(\rho_n - \gamma_n)I(K_{\alpha,n} - \delta \leq L_n < K_{\alpha,n})] \\ &\geq \delta E_0(\gamma_n I(L_n < K_{\alpha,n} - \delta)) - \delta E_0(\gamma_n I(K_{\alpha,n} - \delta \leq L_n < K_{\alpha,n})) \\ &= \delta P_{0,n}(\rho_n \neq \gamma_n \wedge L_n < K_{\alpha,n} - \delta) - \delta P_{0,n}(K_{\alpha,n} - \delta \leq L_n < K_{\alpha,n}). \end{aligned}$$

Since the last probability can be made arbitrarily small by letting  $\delta$  approach zero, whereas by (A4) the first probability in the last equality remains bounded away from zero, it follows that

$\liminf_{n \rightarrow \infty} (E_1(\rho_n) - E_1(\gamma_n)) > 0$ . This result carries over to the general case where  $\rho_n$  is replaced by the asymptotically equivalent test  $\tau_n$  and the exact  $\alpha$ -level test  $\gamma_n$  is replaced by an asymptotic  $\alpha$ -level test. Denoting  $\alpha_n = E_0(\gamma_n)$ , where  $\alpha_n \rightarrow \alpha$ , equality (A2) then only holds in the limit,  $\alpha$  in (A6) needs to be replaced by  $\alpha_n$ , and consequently inequality (A7) now holds in "liminf", which is just fine.

PROOF OF LEMMA 5: The conditions of Lemma 5 imply those of Lemmas 3 and 4.

PROOF OF LEMMA 6: Define  $m(N) = \inf\{n: P[|A_{n,N} - B_N| > 1/N] < 1/N\}$  and let  $N_n^* = \max\{N: m(N) \leq n\}$ .  $N_n^*$  is monotonic and cannot be bounded as otherwise there would exist a constant  $c$  such that  $N_n^* \leq c$  for all  $n$ , therefore  $m(c+1) > n$  for all  $n$ . Thus the existence of a sequence  $N_n$  converging to infinity and bounded by  $N_n^*$  is guaranteed. Then for all  $\varepsilon > 0$ ,  $P(|A_{n,N_n}| > \varepsilon) \leq P(|A_{n,N_n} - B_{N_n}| > \varepsilon) + P(|B_{N_n}| > \varepsilon)$ . Clearly, the right-hand side converges to zero.

PROOF OF LEMMA 7: Let  $x_i = \sum_{j=1}^n y_j^2$ . Then  $\sum_{j=1}^n x_j = 1$  implies  $\sum_{j=1}^n j y_j^2 = 1$ . The L.P. problem involved can now be put in a Lagrange framework, with Lagrange function  $L(y_1, \dots, y_n, \mu) = \sum_{j=1}^n c_j \sum_{i=j}^n y_i^2 + \mu(1 - \sum_{j=1}^n j y_j^2)$ . The solution involved follows now easily from the first-order conditions, in particular the condition  $(\partial/\partial y_k)L(y_1, \dots, y_n, \mu) = 2k y_k [(1/k) \sum_{j=1}^k c_j - \mu] = 0$

## REFERENCES:

ANDREWS, D.W.K. and W.PLOBERGER (1993): "Admissibility of the Likelihood Ratio Test When a Nuisance Parameter is Present Only Under the Alternative", Cowles Foundation Discussion Paper No. 1058, Yale University

ANDREWS, D.W.K. and W.PLOBERGER (1994): "Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative", *Econometrica*, 62, 1383-1414.

BIERENS, H.J. (1982): "Consistent Model Specification Tests", *Journal of Econometrics*, 20, 105-134

- BIERENS,H.J. (1984): "Model Specification Testing of Time Series Regressions", *Journal of Econometrics*, 26, 323-353
- BIERENS,H.J. (1987): "ARMAX Model Specification Testing, with an Application to Unemployment in the Netherlands", *Journal of Econometrics*, 35, 161-190
- BIERENS,H.J. (1988): "ARMA Memory Index Modeling of Economic Time Series (with discussion)", *Econometric Theory*, 4, 35-59
- BIERENS,H.J. (1990): "A Consistent Conditional Moment Test of Functional Form", *Econometrica*, 58, 1443-1458
- BIERENS,H.J. (1994a): "Comment on: 'Artificial Neural Networks: An Econometric Perspective'", *Econometric Reviews*, 13, 93-97
- BIERENS,H.J. (1994b): *Topics in Advanced Econometrics: Estimation, Testing, and Specification of Cross-Section and Time Series Models*, Cambridge: Cambridge University Press
- BIERENS,H.J. and J.HARTOG (1988): "Nonlinear Regression with Discrete Explanatory Variables, with an Application to the Earnings Function", *Journal of Econometrics*, 38, 269-299
- BILLINGSLEY,P. (1968): *Convergence of Probability Measures*, New York: John Wiley
- CHOW,Y.S. and H.TEICHER (1988): *Probability Theory: Independence, Inter-changeability, Martingales*, New York: Springer
- DE JONG,R.M. (1995): "On the Bierens Test Under Data Dependence", *Journal of Econometrics* (forthcoming)
- DE JONG,R.M. and H.J.BIERENS (1994): "On the Limiting Behavior of a Chi-Square Type Test If the Number of Conditional Moments Tested Approaches Infinity", *Econometric Theory*, 10, 70-90
- DUNFORD,N. and J.T.SCHWARTZ (1963): *Linear Operators, Part II: Spectral Theory*, New York: Wiley Interscience
- ENGLE,R.F., D.F.HENDRY and J-F.RICHARD (1983): "Exogeneity", *Econometrica*, 51, 277-304
- GOZALO,P.L. (1993): "A Consistent Model Specification Test for Nonparametric Estimation of Regression Function Models", *Econometric Theory*, 9, 451-477
- HANSEN,B.E. (1990): "Inference When a Nuisance Parameter is Not Identified Under the Null Hypothesis", Rochester Center for Economic Research Working Paper 296, University of

Rochester.

HARDLE,W. and E.MAMMEN (1993): "Comparing Nonparametric Versus Parametric Regression Fits", *Annals of Statistics*, 21, 1926-1947

KUNITA,H. (1990): *Stochastic Flows and Stochastic Differential Equations*, Cambridge: Cambridge University Press

McLEISH,D.L. (1974): "Dependent Central Limit Theorems and Invariance Principles", *Annals of Probability*, 2, 620-628

LEE,T.H., H.WHITE and C.W.J.GRANGER (1993): "Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests", *Journal of Econometrics* 56, 269-290

NEWAY,W.K. (1985): "Maximum Likelihood Specification Testing and Conditional Moment Tests", *Econometrica*, 53, 1047-1070

STINCHCOMBE,M.B. and H.WHITE (1991): "Consistent Specification Tests Using Duality", Mimeo, Department of Economics, University of California, San Diego.

STRASSER,H. (1985): *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory*. New York: De Gruyter.

TAUCHEN,G. (1985): "Diagnostic Testing and Evaluation of Maximum Likelihood Models", *Journal of Econometrics*, 30, 415-443

YATCHEW,A.J. (1992): "Nonparametric Regression Tests Based on Least Squares", *Econometric Theory*, 8, 435-451

WHITE,H. (1989): "An Additional Hidden Unit Test for Neglected Nonlinearity in Multilayer Feedforward Networks", *Proceedings of the International Joint Conference on Neural Networks*, Washington, D.C., San Diego: SOS Printing, II: 451-455

WOOLDRIDGE,J.M. (1992): "A Test for Functional Form Against Nonparametric Alternatives", *Econometric Theory*, 8, 452-475

**Footnotes:**

- 1) The thoughtful comments of two anonymous referees, leading to significant improvements of this paper, are gratefully acknowledged.
- 2) A previous version of this paper, entitled "Asymptotic Optimality and Size of the Integrated Consistent Conditional Moment Test of Functional Form" has been presented at the Econometric Society European Meeting 1993, Uppsala, Sweden. The present version has been presented at Cornell University, North Carolina State University, and Humboldt-Universität zu Berlin.
- 3) The financial support from the "Ausseninstitut der Technische Universität Wien" is gratefully acknowledged. The present version of this paper has been presented at the Econometric Society European Meeting 1994, Maastricht, the Netherlands.
- 4) Corollary 1 of Bierens (1990) with  $\exp(u)$  replaced by  $1/(1+\exp(-u))$  provides a proof of why neural network methods work. See Bierens (1994a).